

Einstellungsmessung in Umfragen: Kategorische vs. Magnitude-Skalen

Wegener, Bernd

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Wegener, B. (1978). Einstellungsmessung in Umfragen: Kategorische vs. Magnitude-Skalen. *ZUMA Nachrichten*, 2(3), 3-27. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-210722>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

EINSTELLUNGSMESSUNG IN UMFRAGEN:

KATEGORISCHE VS. MAGNITUDE-SKALEN⁺

Die direkte und eindimensionale Messung von Empfindungsintensitäten bedient sich in der Regel kategorischer Methoden: Die Befragten werden dabei aufgefordert, die zu skalierenden Stimuli einer begrenzten Anzahl von Kategorien zuzuordnen. Bezogen auf eine Menge von Aussagen ist zum Beispiel die Sequenz "trifft voll und ganz zu / trifft eher zu / trifft eher nicht zu / trifft überhaupt nicht zu" ein gebräuchliches Reaktionsspektrum, das dem Befragten vorgegeben wird und innerhalb dessen er sich für eine Alternative entscheiden muß. Manchmal werden nur die Endkategorien einer solchen Antwortskala ausdrücklich benannt, oder man verzichtet überhaupt auf verbale Beschreibungen und gibt die Kategorien als Folge ganzer Zahlen vor, 1 bis 9 z. B., wobei "1" die kleinste, "9" die größte Ausprägung repräsentieren sollen. Die älteste Kategorienskala, die wir kennen, stammt bereits aus dem zweiten vorchristlichen Jahrhundert, die 6-Punkte-Skala des HIPPARCHUS zur subjektiven Helligkeitsbestimmung von Himmelsgestirnen. Verstärkt im 19. Jahrhundert fanden kategorische Skalierungen sowohl in der Psychophysik als auch in den empirischen Sozialwissenschaften Verwendung. Heute sind sie aus der Forschung nicht mehr wegzudenken. DAWES (1972: 96) stellte bei einer Auszählung der empirischen Arbeiten einer wichtigen Fachzeitschrift fest, daß in ca. 60% der Arbeiten eines Jahrgangs zur Messung abhängiger Variablen kategorische Methoden benutzt wurden. Dieser

⁺ Seit etwa zwei Jahren beschäftigt sich ein Forschungsprojekt im Rahmen der Arbeiten zur Methodenentwicklung bei ZUMA mit den Möglichkeiten, Verfahren der psychophysischen Skalierung auf die Messung von Einstellungen in Befragungen zu übertragen. Die bisherigen Ergebnisse sind vielversprechend. Insbesondere die sog. Magnitude-Skalierung scheint als ernsthafte Alternative für herkömmliche Methoden in Frage zu kommen. Um dem wachsenden Interesse an der Anwendung dieses Meßverfahrens entgegen zu kommen, geben die nachstehenden Ausführungen eine an den Nicht-Fachmann gerichtete Einführung in den Problembereich. Ausführliche Forschungsberichte sind in Vorbereitung. Im Frühjahr 1979 wird außerdem der Bericht über das internationale Symposium zur Sozialen Psychophysik erscheinen, das im Oktober 1978 bei ZUMA stattfand (s. ZUMANACHRICHTEN 2: 47f). Das Forschungsprojekt zur psychophysischen Skalierung wird bei ZUMA von Bernd Wegener betreut, der auch Verfasser dieses Artikels ist.

Prozentsatz dürfte für die Umfrageforschung bei weitem höher anzusetzen sein. Einstellungsmessung in Umfragen ist nachgerade immer kategorisches Messen.

Diese große Verbreitung kategorischer Meßmethoden steht allerdings in Widerspruch zu ihrer methodischen Qualität.

- 1) Kategoriaskalen sind, erstens, praktisch immer Ordinalskalen. Die Zulässigkeit statistischer Operationen ist damit beträchtlich eingeschränkt. Es ist allerdings eine verbreitete Praxis, den Kategorien einer solchen Skala reelle Zahlen in aufsteigender oder absteigender Richtung zuzuweisen und sie unbesehen im Sinne von Intervallmessungen zu behandeln. Im Hinblick auf die übliche Aggregation über Individuen, die Berechnung arithmetischer Mittelwerte und Standardabweichungen nennen LUCE & GALANTER (1963:264f) das zurückhaltend "a bit hair-raising".
- 2) Ein zweites, damit eng verflochtenes Problem stellt die verbale Benennung kategorischer Positionen dar. Adjektivische oder adverbelle "labels" beeinträchtigen die Gleichabständigkeit und Ordnung der Kategorien, weil ihre individuelle semantische Interpretation unkontrolliert bleibt. Aber auch die Beschränkung auf numerische Benennungen löst das Problem in keiner Weise. Jede alternative monotone Zahlenfolge kommt hierfür in Frage und läßt funktionale Beziehungen beliebig werden.
- 3) Bei kategorischen Methoden besteht, drittens, die Gefahr des Informationsverlustes. Einzuschätzende Objekte, die für den Betrachter in bezug auf die relevante Dimension ähnlich sind oder eng beieinanderliegen, erhalten ein und dieselbe Kategorienzuordnung. Das heißt, daß sie selbst dann, wenn das Subjekt zwischen ihnen noch differenzieren kann, in der Skalierung ununterscheidbar werden. Unter inhaltlichen Aspekten ist dieser Verzicht auf Information oft nicht zu vertreten und kann sogar zu Fehlplatzierungen führen.
- 4) Praktisch vielleicht das gravierendste Problem besteht in der zwangsläufigen Begrenzung des Antwortspektrums auf eine vorgegebene Anzahl von Kategorien. Die Breite möglicher Reaktionen von Subjekten wird damit künstlich gestaucht oder auseinandergezogen. Interpersonelle Vergleiche, zumal in den Endkategorien, sind damit bedeutungslos.

5) Schließlich verfügen Kategoriaskalen über keine eingebauten Gütekriterien, an denen sich die Qualität einer Messung ausweisen ließe. Unter den konkreten Gegebenheiten der Feldforschung sind sogar Reliabilitäten schwer abzuschätzen, weil Wiederholungsmessungen in der Regel nicht vorliegen und testtheoretische Halbierungsverfahren nicht anwendbar sind.

Angeichts dieser Liste von Nachteilen ist der Einsatz besserer Methoden der Einstellungs- und Überzeugungsmessung geboten. Freilich gibt es andere psychologische Skalierungsverfahren (TORGERSON 1958, DAWES 1972), die meisten von ihnen sind allerdings aufwendig und lassen sich nur in künstlichen Experimentalsituationen realisieren. Sie setzen z.B. die Erhebung vollständiger Paarvergleichs- oder Ähnlichkeitsmatrizen, umständliche Rangordnungsaufgaben oder Sortings voraus. Wo Daten hingegen nur mit dem Instrument der direkten Befragung in Interviews gesammelt werden können, fehlt bislang eine Alternative zum kategorischen Messen.

Es scheint aber, daß eine Änderung dieser Situation in Aussicht steht. Die moderne Psychophysik (MARKS 1974, STEVENS 1975) kann ein Skalierungsverfahren anbieten, das als Ersatz für das kategorische Messen von Einstellungen unter Umfragebedingungen in Frage kommt. Bereits Ende der sechziger Jahre wurden Versuche unternommen, die Methoden der direkten numerischen Größenschätzung (magnitude estimation) von der Skalierung subjektiv empfundener physikalischer Reizintensitäten auf die Messung politischer, sozialer und anderer Einstellungsvariablen zu übertragen (u.a. SHINN 1969, HAMBLIN 1974). Untersuchungen der letzten Jahre zeigen darüber hinaus, daß die Methoden ohne weiteres auch in unterschiedlichen Bereichen der Feldforschung anwendbar sind (LODGE et al. 1975, 1976a, 1976b). Die Vorteile dieser Möglichkeit werden von inhaltlich forschenden Sozialwissenschaftlern allerdings erst zögernd genutzt (MURPHY & TANENHAUS 1974, LODGE et al. 1976b, BECK, BRATER & WEGENER 1979).

Ein Projekt im Rahmen der ZUMA-Arbeiten zur Methodenentwicklung möchte zur größeren Verbreitung dieser Verfahren beitragen und beschäftigt sich mit deren Ausarbeitung und Prüfung im deutschen Sprachraum. Die nachfolgenden Ausführungen skizzieren die bisher geleistete Arbeit. Zunächst wird die Theorie der Magnitude-messung dargestellt (1), das äußere Vorgehen bei der Messung im Umfragekontext

wird beschrieben (2), und schließlich kommen einige Resultate von Skalenkonfrontationsstudien zur Sprache (3).

1) Zur Psychophysischen Theorie

Drei Arten des Messens

Es ist üblich geworden, in der Psychophysik drei Arten der Empfindungsmessung zu unterscheiden (STEVENS 1957): Die sog. poikilitische Messung, die Teilungsmessung und die Magnitudemessung.

Das poikilitische Messen ($\pi\omicron\kappa\iota\lambda\lambda\omega$ = künstlich bilden) wurde vor allem durch Gustav FECHNER (1860) erforscht. FECHNER unterstellte, daß sich eine Empfindungsskala durch die Addition sog. ebenwahrnehmbarer Empfindungsunterschiede oder "just noticeable differences" (JND) aufbauen ließe. Das Verfahren der Messung setzt also voraus, daß man Unterschiedsschwellen bestimmt und sich eine aktuelle Empfindungsstärke als durch gleiche JNDs zusammengesetzt denkt. Insofern die Operationalisierung der Unterschiedsschwellen bestimmte statistische Annahmen machen muß, basieren JND-Skalen auf Verteilungsannahmen in bezug auf Konfusionsmatrizen. Das FULLERTON-CATTELL-Theorem (1892) bzw. das "law of comparative judgment" THURSTONEs (THURSTONE & CHAVE 1929) stellen die Weiterentwicklung dieser Methoden für nicht-physikalisches Stimulusmaterial dar.

Die Teilungsmessung verlangt, daß der Beobachter sein Empfindungskontinuum in gleichabständige Intervalle unterteilt und Stimulusgrößen diesen Intervallen zuordnet. Beispiele sind alle Formen der Differenzenmessung, aber vor allem auch das kategorische Messen.

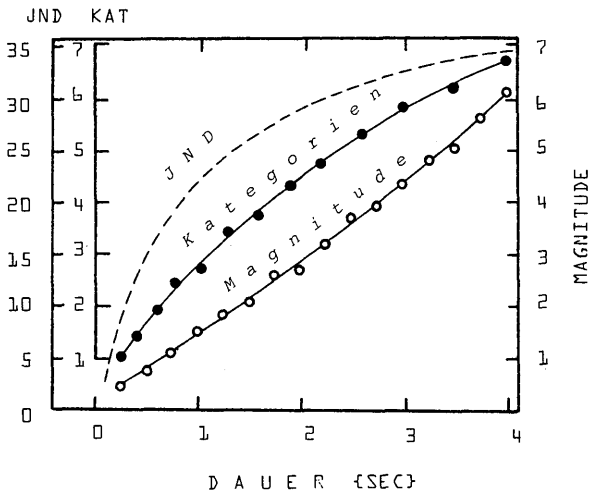
Die Magnitudemessung, schließlich, unterstellt, daß ein Beobachter unmittelbar die Größe einer Empfindungsintensität angeben kann. Er tut dies z.B., wenn er Empfindungsstärken im Kontinuum der reellen Zahlen ausdrückt. Insofern die Eindeutigkeit von Verhältnissen eine der kennzeichnenden Eigenschaften des Relativs der reellen Zahlen ist, stellen sich Größenschätzungen als indirekte relative Verhältnisschätzungen dar. Damit ist zwar noch nicht sichergestellt, daß die Resultate dieser impliziten Verhältnisschätzungen auch die formale Definition von

ZUMA

Verhältnisskalen erfüllen. Als Leistungen werden von Versuchspersonen bei Magnitudeskalierungen aber Verhältnisschätzungen verlangt: Wenn A und B zwei sukzessiv dargebotene Stimuli sind, und B löst bei der Versuchsperson eine z.B. doppelt so starke Empfindung als A aus, dann wird erwartet, daß die Versuchsperson den numerischen Betrag, den sie A zugeordnet hat, bei der Einschätzung von B verdoppelt, usw. Bei der Wahl der ersten Zahl (für A) ist die Versuchsperson frei; jede positive, von Null verschiedene Zahl kommt hierfür in Betracht. Die nachfolgenden Stimuli hingegen müssen mit Zahlen versehen werden, die auf diese Nennung bezogen sind und sie drücken damit - im Zahlenkontinuum - die relativen Verhältnisse aus, in denen die Empfindungsstärken zueinander stehen.

Abb. 1 gibt die Resultate eines typischen psychophysischen Experiments und die psychophysischen Funktionen, die man hier findet, wieder (STEVENS 1957). Einzuschätzen nach den drei unterschiedlichen Skalierungsverfahren war in diesem Fall die Dauer von dargebotenen Tonstimuli. Die JND-Skala erweist sich dabei annähernd als Logarithmus der physikalischen Intensität, die Magnitudeskala als Potenzfunktion, und die kategorische Skala ist in der Regel eine konkave Funktion, die zwischen den beiden anderen liegt. Alle drei Skalen sind Empfindungsskalen, aber welche ist die "wahre"?

Abb. 1



ZUMA

Zunächst wird man sagen müssen, daß die drei Skalen als Operationalisierung von Empfindungsstärken gleichberechtigt sind. Über die lange Geschichte der Psychophysik jedoch haben sich zunehmend die Vorzüge der Magnitudeskalierung herausgestellt, und zwar insbesondere unter dem Gesichtspunkt der Möglichkeit zu einer allgemeinen psychophysischen Theorie.

Die wesentlichsten Argumente gegen die JND-Skala sind vor allem, daß

1. die Konstanz der ebenwahrnehmbaren Unterschiede über die ganze Breite eines Empfindungskontinuums nicht gegeben ist. Vielmehr wird die Größe der Unterschiedswahrnehmungen mit der Reizgröße größer (EKMAN & KÜNNAPAS 1957, BJÖRKMAN 1960, auch bereits BRENTANO 1974:I, 96 ff und JAMES 1892).
2. Ein weiterer - historisch bedeutsamer - Umstand ist, daß die Konstruktion der JND-Skalen intuitiver Alltagserfahrung widerspricht. Die Konstruktion von (logarithmischen) db-Skalen in der Akustik z.B. spiegelt nicht die wirklichen Empfindungsrelationen wider, insofern etwa ein Motoren-geräusch von 80 db nicht doppelt so laut wie eines von 40 db, sondern als entschieden lauter empfunden wird (CHURCHER 1935). Gegen die Kategorialskala andererseits spricht neben den oben aufgeführten Nachteilen u.a., daß sie nicht stabil ist, d.h., daß sie von einer Fülle von Kontexteffekten abhängt: Anzahl der Kategorien, Breite des Reizkontinuums, Benennung der Kategorien, Abstände der Reize auf dem physikalischen Kontinuum, Anker-effekte usw. Außerdem weiß man inzwischen, daß kategorisches Messen von individuellen Antwortstilen beeinflusst wird, die nicht eindeutig zu systematisieren sind, da sie offenbar Ausdruck einer Vielzahl experimenteller Parameter sind und nicht zuletzt Erwartungen, Wertungen und Vorerfahrungen der Befragten widerspiegeln (cf. z.B. EISER & STROEBE 1972).

Der Vorzug der Magnitudeskalen hingegen besteht zu allererst in ihrer enormen Stabilität: Die psychophysischen Funktionen sind zuverlässig als Potenzfunktionen replizierbar. Die Stabilität der Magnitudeskalen ist zudem spezifisch: Im Beispiel aus Abb. 1 läßt sich die psychophysische Funktion zur

objektiven Zeitdauer als Potenzfunktion mit dem Exponenten 1.1 ausdrücken. In etwa dasselbe gilt z.B. für die Einschätzungen von räumlicher Länge; hier finden sich psychophysische Funktionen mit einem Exponenten von 1.0. Die Skalierung von anderen physikalischen Modalitäten (Helligkeit, Lautstärke, Gewicht, Fläche, Kraft usw.) führt ebenfalls zu Potenzfunktionen, jedoch mit anderen Exponenten (vgl. Tab. 1). Diese Entdeckung hat zur Formulierung des allgemeinen psychophysischen Potenzgesetzes geführt:

$$R = a S^{\beta}, \quad (1)$$

wobei R den Vektor der numerischen Reaktionen im Sinne des magnitude estimation bzw. S den Vektor der physikalischen Reizintensitäten repräsentiert. Nach diesem Gesetz hat 1. jede psychophysische Funktion die Form einer Potenzbeziehung (woraus folgt, daß gleiche Stimulusverhältnisse gleiche Empfindungsverhältnisse hervorrufen), und 2. jede physikalische Modalität ihren charakteristischen Exponenten. Für etwa 30 Modalitäten sind die psychophysischen Funktionen inzwischen erforscht, und "ihr" Exponent ist bekannt.

Tab. 1

| | |
|-------------------------------------|------|
| Helligkeit (innerhalb 5°) | 0.33 |
| Lautheit (3000 Hz) | 0.67 |
| Visuelle Fläche | 0.70 |
| Geschmack (Saccharin) | 0.80 |
| Kälte (Hautkontakt) | 1.00 |
| Räumliche Länge (Linien) | 1.00 |
| Zeitdauer (Weißes Geräusch) | 1.10 |
| Gewicht | 1.45 |
| Muskelkraft | 1.70 |
| Elektrischer Schock | 3.50 |

Cross-modality matching

Die Form des Potenzgesetzes ist nun nicht auf die Einschätzung mit Hilfe des Zahlenkontinuums begrenzt. Es ist ja auch möglich, die Intensität einer

Helligkeitsempfindung z.B. mit räumlicher Länge, die manipulierbar ist, auszudrücken, oder durch die Einstellung unterschiedlich lauter Töne. In dem Fall würde die Instruktion lauten, einen Ton so einzustellen, daß er so laut für die Versuchsperson ist wie das dargebotene Licht hell, und dies über eine Serie von Helligkeitsreizen.

In bezug auf ein solches "cross-modality matching" (CMM) läßt sich die zu erwartende Funktion zwischen den beiden involvierten Modalitäten vorher-sagen: Wenn für beide Modalitäten gesicherte Werte über die Exponenten der numerischen Skalierungen vorliegen, so daß

$$R_1 = a_1 S_1^{\beta_1},$$

$$R_2 = a_2 S_2^{\beta_2},$$

dann ergibt sich für die Einschätzung einer der beiden Modalitäten durch die andere (wegen $R_1 = R_2$)

$$S_1 = (a_2/a_1) S_2^{\beta_2/\beta_1}. \quad (2)$$

D.h., nicht nur die Form der Beziehung als Potenzfunktion, sondern auch deren Exponent ist prognostizierbar. Empirisch sind diese CMM-Funktionen mit großer Zuverlässigkeit zu bestätigen (J.C. STEVENS et al. 1960, STEVENS 1966a, J.C. STEVENS & MARKS 1965), woraus die Validität der Magnitude-schätzungen im Zahlenkontinuum folgt, bzw. in reziproker Argumentation: Es scheint, daß die Idee von "Größe" ein universeller Bestandteil unseres Urteilsvermögens ist und eine "Rationalität" aufweist, die keinesfalls nur auf den Umgang mit reellen Zahlen beschränkt ist.

Indirektes cross-modality matching

Das CMM-Paradigma und die bewährten psychophysischen Relationen lassen sich auch zur Skalierung nicht-metrischer Propositionen benutzen, z.B. von Sympathie, Präferenz, Ausmaß von Zustimmung, Zufriedenheit usw. (STEVENS 1966b). Dabei wird eine Liste der zu skalierenden Items dargeboten (z.B. Parteinamen), und der Befragte stellt ein oder produziert in mindestens zwei Modalitäten die physikalischen Intensitäten, die das relative Verhältnis

der von ihm erlebten Empfindungsstärken in bezug auf die Items (etwa unter dem Gesichtspunkt von Präferenz) ausdrücken. Wenn man hierzu Modalitäten benutzt, deren theoretische Exponenten bekannt sind, dann ist die Relation

$$M_1 = a M_2^{\beta_2 / \beta_1} \quad (3)$$

vorhersagbar. (M_1 und M_2 sind die Vektoren der eingestellten Werte in den beiden Modalitäten.) Dieses Vorgehen bezeichnet man als das indirekte cross-modality matching (ICMM), weil die Stimuli selbst - für die eine objektive Metrik ja nicht vorliegt - hierbei nicht in Erscheinung treten, sondern nur die zweifachen Reaktionen auf sie. Wenn diese in dem angegebenen Verhältnis stehen, kann man schließen, daß die Skalierung im Sinne des psychophysischen Modells erfolgreich war. Weiterhin lassen sich Skalenwerte für die Items i durch geometrische Mittlung der transformierten Werte beider Modalitäten berechnen¹⁾:

$$S_i = (M_{1i}^{\beta_1} M_{2i}^{\beta_2})^{1/2}. \quad (4)$$

Die Prüfung, ob eine signifikante Abweichung des empirischen ICMM-Exponenten vom theoretischen vorliegt oder nicht, kann regressionsanalytisch erfolgen, weil (3) durch Logarithmieren linearisiert werden kann:

$$\log M_1 = (\beta_2 / \beta_1) \log M_2 + \log a. \quad (5)$$

Die Steigung dieser linearen Gleichung drückt den theoretischen Exponenten aus. Bei der Berechnung seines empirischen Wertes ist zu beachten, daß beide beteiligten Variablen, M_1 und M_2 , abhängige Meßreihen sind, so daß zur Regressionsberechnung ein Modell, das Fehlerkomponenten in beiden Variablen berücksichtigt, anzuwenden ist (MADANSKY 1959, CROSS 1974).

1)

Der Gebrauch geometrischer Mittelwerte als Aggregierungsstrategie hat sich praktisch als brauchbar erwiesen, weil der Einfluß fehlerhafter Extremwerte minimiert wird (STEVENS 1975: 269ff, INDOW & STEVENS 1966). Darüber hinaus kann gezeigt werden (J.C. STEVENS 1957), daß die Verteilung der Logarithmen der Magnitudewerte angenähert normal ist, woraus sich die multiplikative Mittlung als das angemessene Zentralmaß ergibt. Auf der Basis dieser empirischen Befunde ist das psychophysische Modell, wie es hier als Nukleus dargestellt wird und sich in den Gleichungen (1) und (2) manifestiert, multivariat weiterentwickelt worden (CROSS 1974, 1978), indem für replikative Messungen alle Varianz in M einer multiplikativen und lognormal verteilten Zufallsvariablen zugeschrieben wird.

SCALING INFORMATION - BERUFPRESTIGE

ALL SUBJECTS

----LVAR = 1 THE "STANDARD" STIMULUS IS THE 1 ST STIMULUS,

EXPONENT = 1,0000 FOR MODALITY 1 PRESTIGE LP
EXPONENT = 1,0000 FOR MODALITY 2 PRESTIGE ME

NOTE THAT THE STIMULUS NAMES AND VARIABLE NUMBERS USED BELOW ARE THOSE DEFINED FOR MODALITY 1

| | | | ABSOLUTE SCALE VALUE | RELATIVE SCALE VALUE | LOG. OF REL. SCALE VALUE |
|-------------------|----|---------|-------------------------|-------------------------|-----------------------------|
| RECHTSANW.GEHILFE | LP | (V 319) | 34,03 | 50,00 | 1,699 |
| SCHREINER | LP | (V 320) | 24,97 | 36,69 | 1,565 |
| ELEKTROINST. | LP | (V 321) | 33,64 | 49,42 | 1,694 |
| CHEMIELABORANT | LP | (V 322) | 37,38 | 54,92 | 1,740 |
| EINZELHANDELSKFM. | LP | (V 323) | 36,93 | 54,25 | 1,734 |
| GAERTNER | LP | (V 324) | 19,16 | 28,15 | 1,450 |
| KELLNER | LP | (V 325) | 18,40 | 27,03 | 1,432 |
| KFZ-MECHANIKER | LP | (V 326) | 35,13 | 51,62 | 1,713 |
| BUCHHAEENDLER | LP | (V 327) | 26,05 | 38,27 | 1,583 |
| MAURER | LP | (V 328) | 20,52 | 30,15 | 1,479 |
| METZGER | LP | (V 329) | 25,12 | 36,91 | 1,567 |
| BUCHDRUCKER | LP | (V 330) | 24,29 | 35,69 | 1,553 |
| TECHN. ZEICHNER | LP | (V 331) | 40,11 | 58,93 | 1,770 |

REGRESSION ANALYSIS - BERUFPRESTIGE

ALL SUBJECTS

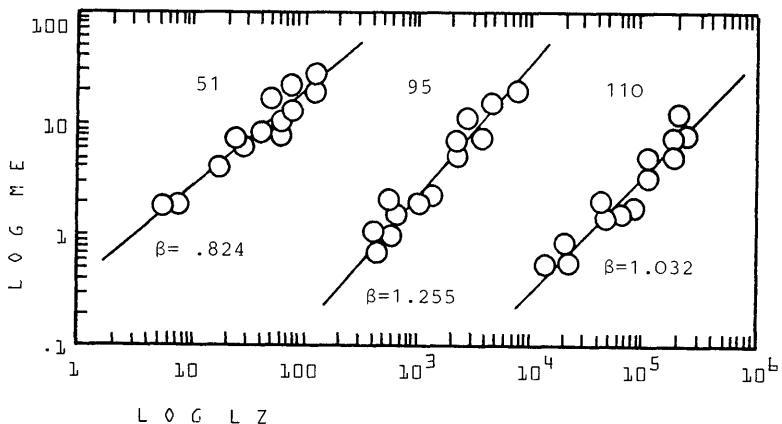
----LVAR = 1 "X" = MODALITY 1 PRESTIGE LP
"Y" = MODALITY 2 PRESTIGE ME

| | | | | | |
|------------------------------------|---|------------|------------------------------------|---|------------|
| MEAN X | = | 1,716 | MEAN Y | = | 1,177 |
| STANDARD DEVIATION X | = | 0,1183 | STANDARD DEVIATION Y | = | 0,1183 |
| SLOPE (X INDEPENDENT, Y DEPENDENT) | = | 0,9764 | SLOPE (Y INDEPENDENT, X DEPENDENT) | = | 0,9756 |
| INTERCEPT (X INDEP., Y DEPENDENT) | = | 0,3178 | INTERCEPT (Y INDEP., X DEPENDENT) | = | 3,691 |
| VARIANCE ABOUT REGR. (Y DEPENDENT) | = | 0,6640E=03 | VARIANCE ABOUT REGR. (X DEPENDENT) | = | 0,6634E=03 |
| CORRELATION COEFFICIENT (R) | = | 0,9760 | SLOPE (BE) = X AND Y DEPENDENT | = | 1,000 |
| R SQUARED | = | 0,9526 | STANDARD ERROR OF SLOPE (BE) | = | 0,2179 |
| F RATIO FOR LINEARITY | = | 220,9 | UPPER LIMIT OF SLOPE (BE) | = | 1,161 |
| T-RATIO | = | 0,1873E=02 | LOWER LIMIT OF SLOPE (BE) | = | 0,8617 |
| INTERCEPT (AE) = X AND Y DEPENDENT | = | 0,2891 | "N" EXCLUDING MISSING DATA | = | 13 |

THE SLOPE (BE) IS NOT SIGNIFICANT GIVEN A USER SUPPLIED TVAL= 2,2010 AND ASSUMING A THEORETICAL EXPONENT= 1,0000

Abb. 2 gibt für drei Befragte einer Untersuchung zur Berufswahl (BECK, BRATER & WEGENER 1979) die Einschätzungen von 13 Lehrberufen in Hinsicht auf ihr Berufsprestige wieder. Bei der Erhebung wurden numerische Magnitudeschätzungen (ME) und Linien (LZ), mit denen die Befragten das Ausmaß des Prestiges angeben sollten, benutzt. Die Abbildung trägt log ME gegen log LZ auf (Koordinatenwerte sind nach unten bzw. nach rechts verschoben). Für alle 122 Befragten der Untersuchung beträgt die mittlere Korrelation zwischen den beiden Modalitäten .942. Die Steigungen der Geraden sind nicht signifikant von 1.0, dem theoretisch zu erwartenden Wert (vgl. Tab. 1), verschieden. Für das aggregierte Sample finden sich die Resultate der Regressionsanalyse der logarithmierten ME- und LZ-Werte in Tab. 2. ("LZ" erscheint im Ausdruck als "LP".) "SLOPE (BE)" ist der Wert des empirischen ICMM-Exponenten, der sich in diesem Fall numerisch vom erwarteten Wert nicht unterscheidet. In Tab. 3 sind die aggregierten Skalenwerte für alle 122 Befragten aufgeführt. Relative Skalenwerte ergeben sich durch Ähnlichkeitstransformationen der absoluten Werte. In diesem Fall wurde für V 319 der Skalenwert 50.00 als "Standard" festgelegt.

Abb. 2



2) Prozeduren

Modalitäten und Formate

Die Anwendung von Magnitudemessungen in Umfragen setzt zunächst die Entscheidung voraus, welche Reaktionsmodalitäten für die Aufgabe angemessen sind und benutzt werden sollen. Es scheint sinnvoll, die folgenden Kriterien bei der Auswahl zu berücksichtigen: 1. Die Modalitäten müssen hinreichend erforscht sein, so daß ihr charakteristischer Exponent β als valide anzusehen ist.

2. Die Kontinuen sollten uneingeschränkt manipulierbar sein, zumindest aber die Herstellung von Endpunktverhältnissen der Reaktionen von 100:1 ermöglichen.

3. Die Modalitäten müssen transportabel und in der Anwendung billig sein.

4. Sie müssen von Interviewern und Befragten problemlos benutzbar sein und,

5., muß ihre Anwendung in konventionelle Interviews integrierbar sein (LODGE et al. 1976b).

Eine Reaktionsmodalität, die diese Bedingungen offenbar ohne weiteres erfüllt, ist die numerische Magnitudeschätzung. Magnitude estimation ist die grundlegende und am häufigsten eingesetzte Methode der direkten Größenskalierung, und man kann plausiblerweise davon ausgehen, daß numerische Magnitudeskalen mit objektiver Zahlengröße ein lineares Verhältnis bilden. Der Exponent für magnitude estimation (ME), mit anderen Worten, ist 1.0.²⁾

Als zweite Reaktionsmodalität, die für Feldarbeiten in Frage kommt, hat sich die Linienmodalität erwiesen: Das Zeichnen von Linien (LZ) proportional zu wahrgenommener Stimulusintensität ist eine gut dokumentierte psychophysische Modalität (STEVENS & GUIRAO 1963, TEGHTSOONIAN & TEGHTSOONIAN 1965), die wie ME von einem Exponenten von 1.0 regiert wird und Reaktionsspannen von 100:1 und mehr

2) Dies ist wahr zumindest im System des psychophysischen Modells. ME stellt das Referenzkontinuum zur Bestimmung der Exponenten für andere Modalitäten dar, wobei sich der Einheitsexponent letztlich definitorisch ergibt. Zur Frage, ob das Zahlenkontinuum in Wirklichkeit nicht als ein "gekrümmtes" wahrgenommen wird, vgl. ATTNEAVE (1962), CURTIS et al. (1968), RULE & CURTIS (1973), SCHNEIDER et al. (1974). Selbst aber wenn sich der Exponent von ME als von 1.0 verschieden erweisen sollte, so würde dies zwar die numerischen Werte der Exponenten aller anderen Modalitäten, nicht aber ihr Verhältnis untereinander beeinflussen.

ermöglicht. Darüber hinaus ist das LZ leicht anwendbar, transportabel, billig, als Papier-und-Bleistift-Verfahren in konventionelle Befragungen integrierbar und es setzt kein besonderes Training der Beteiligten voraus. Am Kriterium der Güte des ICMM-Exponenten hat sich als Papierformat für die Anwendung des LZ in Umfragen das 30-cm-Format als optimal erwiesen (vgl. auch LODGE et al. 1976b)³⁾.

Der Einsatz der direkten Größenschätzung in Umfragen mit ME und LZ erfolgt in zwei Schritten, die als gemeinsamer Block in ein konventionelles Interview eingeschaltet werden.

1. In der eigentlichen Skalierungsphase wird die Befragungsperson aufgefordert, das relative Ausmaß ihrer Einstellungen in Reaktion auf einer Liste von Stimuli (z.B. Statements oder Einzelworte) zuerst mit Linien und dann mit Zahlen auszudrücken. Sie erhält dazu ein Antwortheft, in dem die Stimuli in einem Layout, das ihr genügend Platz für die Reaktionen gibt, aufgeführt sind. Der Interviewer formuliert die dazugehörigen Fragen mündlich.

2. Dieser Skalierungsphase vorgeschaltet ist eine Trainings- und Eichungssequenz. Sie dient neben der Vermittlung des Verfahrens an den Befragten der Bestimmung des sog. Regressionsfehlers, d.i. der individuellen Tendenz der Befragten, die Breite ihrer Antworten zu verkürzen: Niedrige Werte werden über-, hohe unterschätzt. Diese Tendenz zur Mitte ist ein bekanntes Phänomen, das bei allen Formen direkter Skalierung anzutreffen ist (STEVENS & GREENBAUM 1966). Ein Grund für ein Abweichen des empirischen ICMM-Exponenten vom theoretisch zu erwartenden mag entsprechend in der unterschiedlichen Anfälligkeit für Regressionseffekte der beiden Modalitäten zu suchen sein. Wenn diese Abweichung signifikant ist, muß die Skalierung als invalide zurückgewiesen werden. Ein problematischer Fall tritt ein, wenn die Fehlereffekte in beiden Modalitäten symmetrisch auftreten. Der Wert des errechneten ICMM-Exponenten bleibt dann unbeeinflusst, die Skalenwerte jedoch, die

3) Um den Einfluß der Papiergröße zu minimieren, werden die Befragten aufgefordert, mehrere Linien zu zeichnen, wenn sie anders ihre Empfindungsstärke nicht adäquat ausdrücken können. Bei der Kodierung werden die Längen addiert.

nach (4) ermittelt werden, sind u.U. ernsthaft verzerrt. Um den Einfluß des Regressionsfehlers auch unter diesen Bedingungen bestimmen und eliminieren zu können, ist ein Verfahren ausgearbeitet worden, das neben der Einschätzung der Einstellungsstimuli eine Eichung im Gebrauch der benutzen zwei Reaktionsmodalitäten vorsieht: Zunächst wird ein physikalisches Kontinuum, für das der theoretische Exponent bekannt ist, mit Hilfe der beiden Modalitäten eingeschätzt (sinnvoll sind z.B. Linienlängen, Kreisflächen usw.). Man bestimmt an diesem Material den Regressionsfehler als Abweichung der Schätzungen von der zu erwartenden psychophysischen bzw. CMM-Funktion. Wenn bestimmte Konsistenzbedingungen erfüllt sind, läßt sich schließen, daß dasselbe Ausmaß an Fehler auch bei den nichtmetrischen, sozialen Einschätzungen wirksam ist. Da nachgewiesen werden kann (CROSS 1974), daß die fehlerhaften Schätzungen sich im Sinne einer Potenzfunktion zu den wahren Werten verhalten, äußert sich eine Korrektur bei der Berechnung der Skalenwerte dann einfach als Veränderung der Exponenten, die in Gleichung (4) eingehen.

Die Darbietung von Listen von Stimuli (zwischen zwei und etwa 20) bei der mehrmodalen psychophysischen Skalierung stellt die einfachste Form der Darbietung dar. Es sind aber auch Variationen der Methode möglich. Eine indirekte Anwendung der Magnitudemessung ergibt sich, wenn verbale labels (die z.B. Graduierungen von Zustimmung ausdrücken) zunächst mit Hilfe von zwei Modalitäten skaliert werden. Bei der Beantwortung von inhaltlichen Fragen erhält die Befragungsperson dann die Möglichkeit, aus dem Satz dieser labels einen auszusuchen, um seine Einstellung verbal zu kennzeichnen. Dieser zweite Teil der Aufgabe sieht für den Befragten dann zwar äußerlich wie eine Kategorialskalierung aus, die verbalen Bezeichnungen, die ihm zur Verfügung stehen, sind aber durch die vorgeschaltete Magnitudemessung individuell skaliert, wodurch sich nicht-kategoriale Skalen bei der Beantwortung der inhaltlichen Fragen errechnen lassen.

Ein anderes Vorgehen, dessen Einsatz im Fall singulärer Stimuli angezeigt ist, sieht zunächst eine dichotome Entscheidung der Befragungsperson vor: Sie muß angeben, ob sie in Hinsicht auf ein Item eine eher zustimmende oder eine eher ablehnende Haltung einnimmt, ob sie eher zufrieden oder eher unzufrieden ist usw.

Anschließend drückt sie dann das Ausmaß ihrer Zustimmung/Ablehnung oder Zufriedenheit/Unzufriedenheit mit Magnitudeschätzungen aus. Da Magnitudemessung jedoch relative Verhältnisschätzungen verlangt, muß für ihr Urteil ein Bezug etabliert werden. In den meisten Fällen eignet sich hierfür ein fiktiver Stimulus, der einzuschätzen ist, nämlich die Vorstellung eines "Weder-noch". M.a.W., im Fall singulärer Stimuli wird die Befragungsperson zunächst aufgefordert, eine Bezugszahl bzw. eine Bezugslinie zu produzieren, die z.B. einen Standpunkt der Weder-Zustimmung-noch-Ablehnung repräsentieren, um dann relativ dazu das Ausmaß ihrer Zustimmung oder Ablehnung auszudrücken.

Auswertung

Die Analyse mehrmodaler Magnitudeschätzungen erfolgt im wesentlichen in zwei Schritte: 1. Bestimmung des ICMM-Exponenten als Regressionskoeffizient in Gleichung (5) und Überprüfung, ob eine signifikante Abweichung vom theoretischen Wert vorliegt oder nicht. 2. Skalierung der Stimuli nach Gleichung (4), u.U. mit Exponenten, die den Effekt des individuellen Regressionsfehlers berücksichtigen. Mit dem bei ZUMA entwickelten Programm ZUMAPSYCH liegt erstmals ein Auswertungsprogramm vor, das diese Rechenschritte mit flexiblen Optionen und auf gewünschtem Aggregationsniveau vornehmen kann (s. Tab.2 und 3). Das Programm ist OSIRIS-kompatibel, verfügt über verschiedene Optionen für die Skalenausgabe und die Kennwerte, die die Güte der Skalierung in bezug auf den ICMM-Exponenten kennzeichnen, und gestattet die Bearbeitung großer Datensätze (s. ZUMAPACK-Manual).

3) Einige Resultate

Operationalität

Ist Magnitudemessung "besser" als konkurrierende Meßverfahren, und das heißt im Umfragekontext: Ist sie dem kategorischen Messen überlegen? Bevor diese Frage beantwortet und auch nur sinnvoll gestellt werden kann, muß sichergestellt sein, daß das neue Verfahren auch anwendbar ist und in Befragungen, wie wir sie kennen, überhaupt "läuft". In Hinsicht auf experimentelle und quasi-experimentelle

Messungen von Einstellungsvariablen liegen eine Fülle von Erfahrungen diesbezüglich vor (cf. Literaturübersichten bei STEVENS 1975, LODGE et al. 1975, 1976b, DAWSON 1974). Allerdings handelt es sich hier in der Regel um Studien, die nur eine einzige Reaktionsmodalität benutzen oder bei denen die Erhebungssituationen nicht mit der von standardisierten Einzelinterviews vergleichbar sind. Im Rahmen des ZUMA-Forschungsprojekts zur psychophysischen Skalierung wurde eine Reihe von Untersuchungen mit zunehmender Komplexität durchgeführt, die die Verwendung der mehrmodalen direkten Größenskalierung in Einzelbefragungen demonstrieren sollten. Als Kriterium für die Operationalität war immer die Logik des ICMM-Paradigmas ausschlaggebend. In Gruppenuntersuchungen mit Pilotcharakter wurden zunächst die Formate für die Instruktionen, das Layout und die Analyseprozeduren ausgetestet und dann in regulären Interviewsituationen im Rahmen inhaltlicher Projekte ausprobiert. Eine repräsentative Untersuchung, bei der die Mitarbeit der Kooperationsinstitute ZUMAs vorgesehen ist, ist in Vorbereitung.

Man kann uneingeschränkt feststellen, daß die Form der Magnitudemessung, wie sie bei ZUMA nunmehr in Gebrauch ist, bei der Anwendung in Befragungen keinerlei Schwierigkeiten macht. Befragte unterschiedlicher Populationen sind in der Lage, relative Verhältnisschätzungen abzugeben, ebenso wie Interviewer bereits nach kurzer, z.T. schriftlicher Instruktion die Technik im Interview einsetzen können. Darüber hinaus ist das Verfahren populär: Eine Befragung von Interviewern und Befragten ergab, daß die Interviewer im Mittel die Anwendung numerischer Magnitudeschätzungen anderen Abfragetechniken (Zustimmung auf vorgelesene Precodes, Alternativfragen, Kartenspiele) vorziehen, wobei der Umgang mit numerischer Schätzung allerdings gegenüber dem LZ präferiert wird. Die Befragten gaben an, daß ihnen das ME "sehr gut" und das LZ "gut" gefällt (mittlere Rangplätze bei $N = 123$). Dieser Motivierungseffekt ist insbesondere bei langen Interviews hilfreich. Entsprechend ist es auch nicht verwunderlich, daß nur in etwa 1% aller Skalierungen eine Magnitudeskala wegen missing data in beiden Modalitäten nicht konstruiert werden konnte. Befragte, die unfähig waren, die Schätzungen nach der Magnitudemethode anzugeben oder sich weigerten, dies zu tun, sind in unseren bisherigen Untersuchungen mit ca. 1.500 Skalierungen nicht vorgekommen. Im

Gegenteil, 83% der Befragten und 92% der Interviewer geben an, daß sie keinerlei Probleme mit den Verfahren hatten.

Die untersuchten Variablen umfassen neben einigen physikalischen Stimulusreihen (WEGENER 1977) Einstellungen zu politischen Sachverhalten (u.a. Parteienpräferenz, Parteienidentifikation, Links-Rechts-Plazierungen, Wahrnehmung politischer Streitfragen (issues), Parteienkompetenz), zur Berufszufriedenheit, zum Berufsprestige, zu Persönlichkeitsmerkmalen und zu Aussagen zur Berufswahl, sowie eine adjektivische Unterstützungsskala.

Ein wichtiger Posten ist der Zeitaufwand. Schließlich muß bei Magnitudeschätzungen das Stimulusmaterial zweimal durchgegangen werden, und die Befragten müssen die Trainings- und Eichungsphase durchlaufen. Als Faustregel gilt hier, daß bei einer Fragenbatterie von vier Fragen mit durchschnittlich 10 Items pro Frage gegenüber der konventionellen, kategorischen Skalierung ein zusätzlicher Zeitbedarf von 5 - 7 Minuten anzusetzen ist. Dieser Mehraufwand verringert sich relativ zur Gesamtzeit in dem Maße, in dem das psychophysisch zu beantwortende Fragenkontingent umfangreicher wird.

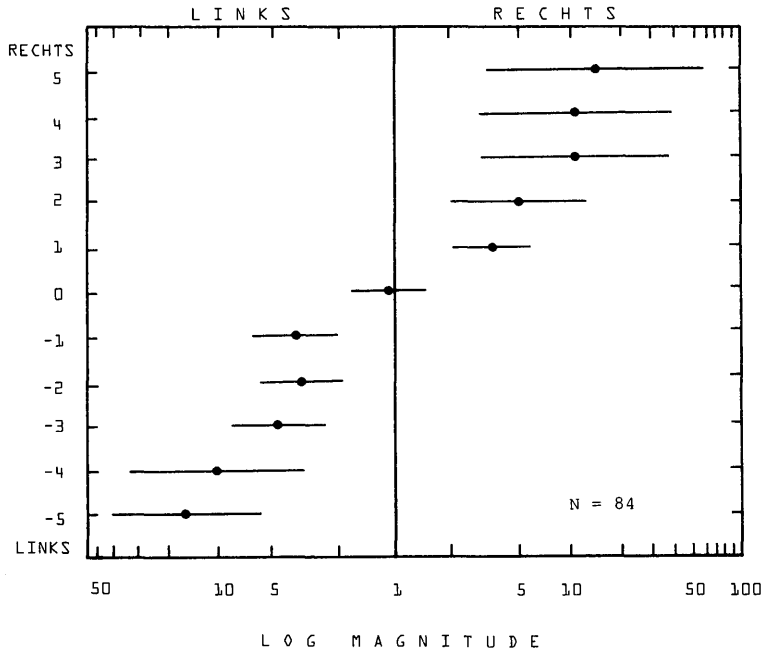
Variabilität

Zwei einschneidende Nachteile der Kategorienskala (s.o.) sind 1. die Begrenztheit der vorgegebenen Antwortalternativen und 2. die Unmöglichkeit, zwischen eng zusammenliegenden Stimuli zu unterscheiden. In fast allen bei ZUMA durchgeführten Untersuchungen sind daher kategorische und Magnitudemessungen parallel durchgeführt worden, um die Auswirkungen dieser Restriktionen sichtbar zu machen. Das Ergebnis all dieser Skalenkonfrontationsstudien ist einheitlich (cf. auch LODGE & FOLEY 1977, LODGE & TURSKY 1978): Die Magnitudemessung ist in der Lage, Variabilität differenzierter als das kategorische Verfahren abzubilden. Wie zu erwarten, zweitens, ist die Streuung in den Endbereichen einer Skala besonders groß, d.h., es werden auch Extremlagen erfaßt, die in der kategorischen Zuordnung mit weniger ausgeprägten Skalenwerten zusammenfallen. Schließlich wird deutlich, daß die Kategorienskalen Fehlplatzierungen begünstigen; es stellt sich nämlich häufig heraus, daß Kategorien im Mittelbereich auf dem Empfindungskontinuum derart eng beieinanderliegen, daß die Befragten zwischen ihnen nicht mehr eindeutig differenzieren können.

ZUMA

Ein typisches Ergebnis ist in Abb. 3 dargestellt. Die Einschätzung nach "links-rechts" in bezug auf vier Parteien, die "ideale" Partei und die eigene Person sind einmal anhand einer bi-polar benannten 11-Punkte-Skala und einmal mit Magnitudeskalierung gemessen und gegeneinander aufgetragen. Dargestellt sind die logarithmischen Mittelwerte aller psychophysischen Einschätzungen pro genannter Kategorie jeweils mit einer logarithmischen Standardabweichung. Neben den angeführten Variabilitätsgesichtspunkten macht die Abbildung deutlich, daß wir es hier im Grund mit zwei verschiedenen Skalen zu tun haben: einer Skala für "links" und einer für "rechts". Beide Teilkurven lassen sich nämlich als Potenzfunktionen rekonstruieren, deren Exponenten unterschiedlich sind. Aufschlußreich ist weiterhin, daß die gesamte Breite der Skala ein Endpunktverhältnis von etwa 120:1 aufweist, während die Kategorienskala nur 11 Abstufungen zur Verfügung hat, um diese enormen und realen Unterschiede abzubilden.

Abb. 3



Kommensurabilität

Sind angesichts dieser Verzerrungen kategorische und Magnitudeskalen überhaupt noch vergleichbar? Können wir noch davon ausgehen, daß beide Skalentypen Identisches messen - eben nur auf unterschiedliche Weise? Im Rahmen der reinen Psychophysik wird diese Frage seit langem intensiv diskutiert, und es finden sich dort Proponenten für beide Standpunkte, für den Standpunkt der grundsätzlichen Verschiedenheit der Skalentypen (z.B. MARKS 1974) und für den Standpunkt ihrer Vergleichbarkeit (KRANTZ 1972). Im letzteren Fall geht es darum, eine Transformationsfunktion zu finden, die kategorische in Magnitudeskalen abbildet und vice versa, so daß, wenn eine solche Funktion existiert, behauptet werden kann, daß ein und dieselbe Empfindung beiden Messungen zugrundeliegt, die lediglich unterschiedlich repräsentiert wird.

Die Entscheidung für einen der beiden Standpunkte wird sicherlich erst verbindlich zu treffen sein, wenn eine vollständige Theorie der Kategorialmessung vorliegt (LUCE & GALANTER 1963:268, MARKS 1968), ein Ziel freilich, das noch in weiter Ferne liegt. Unter Anwendungsgesichtspunkten jedoch ist es sinnvoll, vor der Hand die zweite der beiden Positionen einzunehmen und Aussagen über die funktionale Beziehung der Skalentypen anzustreben. Man kann nämlich nicht erwarten, daß Sozialwissenschaftler zur Übernahme der Magnitudeskalierung motiviert sind, sofern ihnen nicht zumindest grobe Hinweise darüber vorliegen, wie Ergebnisse, die sie mit der neuen Methode zu erwarten haben, sich zu solchen, die sie - vielleicht über Jahre hin - mit konventionellen Verfahren erhalten haben, in Beziehung setzen lassen. Der Gang von "normal science" ist nun einmal kumulativ.

Für Einstellungsvariablen, die unter den besonderen Bedingungen von Umfragen gemessen werden, hat WEGENER (1978) ein additives Potenzmodell für die Vergleichbarkeit von kategorischen und Magnitudeskalen vorgeschlagen, das die Form

$$K + \kappa = a M^{\alpha} \quad (6)$$

hat; wobei K und M die Vektoren der Kategorial- bzw. Magnitudewerte und die übrigen Parameter Konstanten sind. Im Bereich der reinen Psychophysik ist dieses

Modell (MARKS 1968, MARKS & CAIN 1972) neben anderen (EISLER 1962a, 1962b) in der Diskussion. Desungeachtet ist dem Potenzmodell (6) in dem vorliegenden pragmatischen Zusammenhang u.a. deswegen der Vorzug zu geben, weil der Parameter α sich als ein Indikator für sowohl die Anzahl der Kategorien der kategorischen Skala, als auch für die Stimulusbreite erweist: Mit der Erhöhung der Kategorienzahl wird der Exponent α größer, während eine größere perzipierte Stimulusbreite zu kleineren Exponenten bei der Anpassung der beiden Skalentypen führt.

Für eine Reihe von Einstellungsvariablen ließ sich das numerische Verhältnis zwischen kategorischen und Magnitudeskalen im Sinne des additiven Potenzmodells rekonstruieren. Ein iteratives Schätzverfahren zur Optimierung der Werte für κ (ZUMAPSYCH2) führte zu Korrelationen zwischen $\log(K + \kappa)$ und $(\alpha \log M + \log a)$, die im Mittel bei .93, für aggregierte Skalen sogar bei .99 lagen. Dennoch ist aber bei der Interpretation dieser Interskalen- oder Übersetzungsfunktionen Vorsicht geboten: Aus (6) wird deutlich, daß das Anpassungsmodell die kategorischen Skalen als Intervallskalen und die Magnitudeskalen als Log-Intervallskalen behandelt, insofern die an den Rohwerten vorzunehmenden Transformationen die für diese Skalentypen zulässigen sind (STEVENS 1946, 1951). Bis jetzt ist es nur in experimentellen Studien und an physikalischem Stimulusmaterial gelungen, die entsprechenden (algebraischen Differenzen-) Meßstrukturen und gleichzeitig die numerische Beziehung im Sinne von (6) zwischen kategorischen und Magnitudeskalen nachzuweisen (WEGENER 1977, ORTH 1978). Wo dies allerdings nachgewiesen werden kann, gilt, daß die nach (6) resultierende gemeinsame Skala aus K- und M-Skalen die Eindeutigkeit von Verhältnisskalen besitzt (KRANTZ 1972). Die Klärung, ob solche idealen Fälle auch für die Einstellungsmessung in Befragungen realisierbar sind oder nicht, muß dem Fortgang der Forschung überlassen bleiben, in der sich Feld- und Laboratoriumsstudien ergänzen müssen. Im Normalfall ist aber zumindest davon auszugehen, daß Magnitudeskalen das Niveau von logarithmischen Intervallskalen mit $\{ \phi \rightarrow \lambda \phi^{\gamma} \}$ als Klasse zulässiger Transformationen aufweisen.

Theoretische Fruchtbarkeit

Die Vermutung ist nicht unberechtigt, daß die beklagte und ausgewiesen schlechte Korrespondenz zwischen Einstellungsvariablen und Verhalten (z.B. DeFLEUR & WESTIE 1958, CAMPBELL 1963, TITTLE & HILL 1967, SCHWARTZ & TESSLER 1972) wesentlich von der fehlenden Güte der gebräuchlichen Einstellungsskalierung verursacht wird. Jedenfalls können wir nicht genau sagen, was Grund für die enttäuschende Prediktorkraft ist, solange die metrische Qualität von Einstellungsvariablen nicht eindeutig und nach Möglichkeit hoch ist: Falsche Prognosen können ihre Ursachen in der Zugrundelegung unangemessener Prediktoren haben - aber auch in der Schwäche ihrer Messung. Bei dem ausgeprägten Mangel an formalisierten Theorien in vielen Bereichen der Sozialwissenschaften heißt das aber umgekehrt, daß die Güte rivalisierender Methoden zur Einstellungsmessung unter Gesichtspunkten ihrer "theoretischen Fruchtbarkeit" nur schwierig zu bestimmen ist. Die Maximierung von Zusammenhangsmaßen kann alleine nicht ausschlaggeben sein, weil wir nicht wissen, wie groß die Zusammenhänge denn unabhängig von unseren alternativen Operationalisierungen sind.

Aus diesem Grund ist der Methodenvergleich zwischen kategorischer und Magnitudenmessung in erster Linie auf formale Gesichtspunkte angewiesen, und hier deutet alles auf die Überlegenheit der Magnitudeskalierung hin. Gegenüber kategorischen Methoden hat die Magnitudenmessung:

- die Vorzüge eines nonverbalen Verfahrens;
- das Antwortspektrum der Befragten wird nicht durch Vorgaben eingegrenzt;
- die intra- und interindividuelle Varianz wird differenzierter als beim kategorischen Skalieren abgebildet.
- Darüber hinaus liefert die Methode Gütekriterien für die Messung,
- und sie führt zu Skalen mit Log-Intervall-Niveau, ja es besteht sogar die Hoffnung, in Einzelfällen Verhältnisskalen zu erhalten.

Es sind diese Skaleneigenschaften allerdings, die die Frage nach der theoretischen Fruchtbarkeit des Magnitudekonzeptes durchaus beantwortbar machen, wenn auch nur in einem allgemeinen und theorieunspezifischen Sinne: Die Formulierung von funktionalen Beziehungen zwischen verschiedenen Variablen oder zwischen identischen Variablen, die zu unterschiedlichen Zeitpunkten gemessen wurden, ist nur

dort möglich, wo Skaleneigenschaften vorliegen, die isotone und insbesondere affine Transformationen ausschließen. In positiver Formulierung heißt das, daß empirische Gesetzes- und Veränderungsannahmen funktionalen Charakter nur dann annehmen können, wenn die beteiligten Variablen - wie im Fall ihrer Magnitudeskalierung - die Eindeutigkeit von logarithmischen Intervall- oder von Verhältnisskalen aufweisen.

Damit wird hier keinesfalls puristisch der Standpunkt vertreten, daß für die Messung von Einstellungs- und Überzeugungsstärken die Magnitudeskalierung immer und ausschließlich die Methode der Wahl sein sollte. Das Interesse, das man mit einer Messung verfolgt, ist ausschlaggebend. Vom Standpunkt des Methodologen ist darüber hinaus zu betonen, daß Fortschritt in der sozialwissenschaftlichen Skalierungsforschung - in Ermangelung metrischer Validierungskriterien - ganz wesentlich von Vergleichen der Leistungsfähigkeit konkurrierender Meßmethoden und konzeptioneller Paradigmen lebt. Unter der Beschränkung auf die Möglichkeiten der Messung in Umfragen und bei der Gegenüberstellung von kategorischer und Magnitudeskalierung ist es unter diesem Gesichtspunkt nur folgerichtig, in einem Plädoyer für "Magnitude" auch auf die jüngsten Erfolge bei der Untersuchung von Kategoriaskalen zu verweisen, z.B. auf PARDUCCI (1978), ROHRMANN (1978) oder LEVINE & BADEN (1978).

Literatur

- ATTNEAVE, F. Perception and related areas. In: S.KOCH (Ed.).
Psychology: A Study of a science, Vol.IV. New York: McGraw-Hill,
1962, 619-659.
- BECK, U., M.BRATER & B.WEGENER. Berufswahl und Berufszuweisung. Eine empirische Untersuchung zur sozialen Verwandtschaft von Ausbildungsberufen. Frankfurt: Campus, 1979 (im Erscheinen).
- BJÖRKMAN, M. Variability data and direct quantitative judgment for scaling subjective magnitude. Rep. Psychol. Lab., Univ. Stockholm, 1960, 78.
- BRENTANO, F. Psychologie vom empirischen Standpunkt. Leipzig: Duncker & Humblot, 1874 (Neuaufgabe hrsg. von O.KRAUS, Hamburg: Meiner 2/1955, 3 Bde.).
- CAMPBELL, D.T. Social attitudes and other acquired behavioral dispositions. In: S.KOCH (Ed.). Psychology: A study of a science. Vol.VI, New York: McGraw-Hill, 1963, 94-172.
- CHURCHER, B.G. A loudness scale for industrial noise measurement. J. Acoustic.Soc.America, 1935, 6, 216-226.

- CROSS, D. V. Some technical notes on psychophysical scaling. In: H. MOSKOWITZ, B. SCHARF & J. C. STEVENS (Eds.), Sensation and measurement: Papers in honor of S.S. Stevens. Dordrecht: Reidel, 1974, 23-36.
- On judgments of magnitude. Paper presented at International Symposium on Social Psychophysics, Mannheim, October, 9-11, 1978.
- CURTIS, D. W., F. ATTNEAVE & T. L. HARRINGTON. A test of a two stage model of magnitude judgment. Perception & Psychophysics, 1968, 3, 25-31.
- DeFLEUR, M. E. & F. R. WESTIE. Attitude as a scientific concept. Social Forces, 1963, 17-31.
- DAWES R. M. Fundamentals of attitude measurement. New York: Wiley, 1972 (dt.: Grundlagen der Einstellungsmessung. Weinheim: Beltz, 1978).
- DAWSON, W. E. An assessment of ratio scales of opinion produced by sensory-modality matching. In: H. R. MOSKOWITZ, B. SCHARF & J. C. STEVENS (Eds.), Sensation and measurement: Papers in honor of S.S. Stevens. Dordrecht: Reidel, 1974, 23-36.
- EISER, J. R. & STROEBE, W. Categorization and social judgment. London: Academic Press, 1972.
- EISLER, H. On the problems of category scales in psychophysics. Scand. J. Psychol., 1962a, 3, 81-87.
- Empirical test of a model relating magnitude and category scales. Scand. J. Psychol., 1962b, 3, 88-96.
- EKMÁN, G. & T. KÜNNAPAS. Subjective dispersion and the Weber fraction. Rep. Psychol. Lab., Univ. Stockholm, 1957, 41.
- FECHNER, G. Elemente der Psychophysik. Leipzig: Breitkopf und Härtel, 1860.
- FULLERTON, G. S. & J. McK. CATTELL. On the perception of small differences. Philadelphia: Univ. Pennsylvania Press, 1892.
- HAMBLIN, R. L. Social attitudes: Magnitude measurement and theory. In: H. M. BLALOCK (Ed.), Measurement in the social sciences: Theories and strategies. Chicago: Aldine Pbl. Comp., 1974, 61-120.
- INDOW, T. & S. STEVENS. Scaling of saturation and hue. Perception & Psychophysics, 1966, 1, 253-271.
- JAMES, W. Psychology: Briefer course. New York: Holt, 1892
- KRANTZ, D. H. Magnitude estimation and cross-modality matching. J. Math. Psychol., 1972, 9, 168-199.
- LEVINE, L. H. & S. BADEN. The effect of scale size on obtaining information in preference testing. Quality and Quantity, 1978, 12, 169-174.
- LODGE, M., D. V. CROSS, B. TURSKY & J. TANENHAUS. The psychophysical scaling and validation of a political support scale. Am. J. Pol. Sci., 1975, 19, 611-649.
- LODGE, M., D. V. CROSS, B. TURSKY, J. TANENHAUS & R. REEDER. The psychophysical scaling of political support in the 'real world'. Pol. Methodology, 1976a, 2, 3, 159-182.
- LODGE, D., D. V. CROSS, B. TURSKY, M. A. FOLEY & H. FOLEY. The calibration and cross-modal validation of ratio scales of political opinion in survey research. Soc. Sci. Research, 1976b, 5, 352-347.
- LODGE, M. & H. FOLEY. Comparison between category and magnitude scaling of political opinion. Laboratory for Behavioral Research, Stony Brook, 1977.
- LODGE, M. & B. TURSKY. The social-psychophysical scaling of political opinion. Paper presented at International Symposium of Social Psychophysics, Mannheim, October 9-11, 1978.

- LUCE, R.D. & E.GALANTER. Psychophysical scaling. In: R.D.LUCE & E.GALANTER (Eds.), Handbook of mathematical psychology, Vol.I. New York: Wiley, 245-307, 1963.
- MADANSKY, A. The fitting of straight lines when both variables are subject to error. J. Am. Stat. Ass., 1959, 54, 173-205.
- MARKS, L.E. Stimulus-range, number of categories, and the form of the category scale. Am.J.Psychol., 1968, 81, 467-479.
- MARKS, L.E. & W.S.CAIN. Perception of intervals and magnitudes for three prothetic continua. J.Exp.Psychol., 1972, 94, 6-17.
- MARKS, L.E. On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as a science. Perception & Psychophysics, 1974, 16, 358-376.
- MURPHY, W. & J. TANENHAUS. Explaining diffuse support for the United States Supreme Court: An assessment of four models. Notre Dame Lawyer, 1974, 49, 1037-1044.
- ORTH, B. On the scale properties of magnitude estimation and rating scales: A model and some experimental results. Paper presented at International Symposium on Social Psychophysics, Mannheim, October 9-11, 1978.
- PARDUCCI, A. Number of categories: A crucial determinant of the rating scale. Paper presented at International Symposium on Social Psychophysics, Mannheim, October 9-11, 1978.
- ROHRMANN, B. Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. Zeitschr.f.Sozialpsychologie, 1978, 9, 222-245.
- RULE, S.J. & D.W.CURTIS. Conjoint scaling of subjective number and weight. J.Exp.Psychol., 1973, 97, 305-309.
- SCHNEIDER, B., S.PARKER, D.OSTROSKY, D.STEIN & G. KANOW. A scale for the psychological magnitude of number. Perception & Psychophysics, 1974, 16, 43-46.
- SCHWARZ, S. & R. TESSLER. Model for reducing measured attitude behavior discrepancies. J. Person.Soc.Psychol., 1972, 225-236.
- SHINN, A. Jr. The application of psychophysical scaling techniques to measurement of political variables. Institute for Research in Social Science, Univ. North Carolina: Chapel-Hill Press, 1969.
- STEVENS, J.C. A comparison of ratio scales for the loudness of white noise and the brightness of white light. Unpubl.doc.diss., Havard Univ., 1957.
- STEVENS, J.C., J.D. MACK & S.S. STEVENS. Growth of sensation on seven continua as measured by force of handgrip. J. Exp.Psychol., 1960, 59, 60-67.
- STEVENS, J.C. & L.E.MARKS. Cross-modality matching of brightness and loudness. Proc.Nat. Acad.Sci., 1965, 54, 407-411.
- STEVENS, S.S. On the theory of scales of measurement. Science, 1946, 103, 677-680.
- Mathematics, measurement, and psychophysics. In: S.S.STEVENS (Ed.), Handbook of experimental psychology. New York: Wiley, 1951, 1-49.
 - On the psychophysical law. Psychol.Rev., 1957, 64, 153-181.
 - Matching functions between loudness and ten other continua. Perception & Psychophysics, 1966a, 1, 5-8.
 - A metric for the social consensus. Science, 1966b, 151, 530-541.
 - Psychophysics: Introduction to its perceptual, neural, and social prospects. New York: Wiley, 1975.

- STEVENS, S.S. & M. GUIRAO. Subjective scaling of length and area and the matching of length to hardness and brightness. J.Exp.Psychol., 1963, 66, 177-196.
- STEVENS, S.S. & H.B. GREENBAUM. Regression effect in psychophysical judgment. Perception & Psychophysics, 1966, 1, 439-446.
- TEGHTSOONIAN, M & R. TEGHTSOONIAN. Seen and felt length. Psychonom.Soc., 1965, 3, 465-466.
- THURSTONE, L.L. & E.J.CHAVE. The measurement of attitudes. Chicago : Univ. Chicago Press, 1929.
- TITTLE, C. & R.J. HILL. Attitude measurement and prediction of behavior: An evaluation of conditions and measurement techniques. Sociometry, 1967, 30, 199-213.
- TORGERSON, W.S. Theory and method of scaling. New York: Wiley, 1958.
- WEGENER, B. Das Exponentenproblem bei der psychophysischen Skalierung 'sozialer' Variablen. ZUMA-Arbeitsbericht, Mannheim, 1977.
- Fitting category to magnitude scales of opinion. Paper presented at International Symposium on Social Psychophysics, Mannheim, October 9-11, 1978.